DOCUMENT RESUME

ED 163 050 TM 008 090

AUTHOR Steinheiser, Frederick H., Jr.; Hirshfeld, Stephen L.
TITLE Using Stein's Estimator to Predict Universe Scores From Obtained Scores. Research Memorandum 78-19.
INSTITUTION Army Research Inst. for the Behavioral and Social Sciences, Arlington, Va.
PUB DATE Aug 78
NOTE 17p.

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS Bayesian Statistics; Goodness of Fit; Mathematical Models; Measurement; *Prediction; Scores; *Statistical Analysis; *True Scores
IDENTIFIERS *Estimation

ABSTRACT

The scientific implications and practical applications of the Stein estimator approach for estimating true scores from observed scores are of potentially great importance. The conceptual complexity is not much greater than that required for more conventional regression models. The empirical Bayesian aspect allows the examiner to incorporate his/her own degree of prior information about selected examinees. This approach allows for a more accurate estimation of true scores, with the corollary of using fewer test items to achieve those true score estimates. Efron and Morris make the point that there is little penalty for using the rules discussed here because they cannot give larger total mean squared error than the maximum likelihood estimate. This assurance may be a sufficient reason for more careful examination of the utility of the Stein estimator and its limited translation estimators as they apply to behavioral and social science research. A numerical example is given of prediction of a season's batting averages on the basis of a sample of about 10 percent of the times at bat. (Author/CTM)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED163050

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Research Memorandum 78-19

USING STEIN'S ESTIMATOR TO PREDICT UNIVERSE SCORES FROM OBTAINED SCORES

Frederick H. Steinheiser, Jr.,

and

Stephen L. Hirshfeld

UNIT TRAINING AND EVALUATION SYSTEMS TECHNICAL AREA

TM008 090

U. S. Army
Research Institute for the Behavioral and Social Sciences

August 1978

Addendum

A preliminary version of this paper was presented at the meeting of the American Educational Research Association in New York, April 5, 1977.

Erratum

All instances of the "$\int$" (integral sign) should be replaced with "δ" (lower case Greek delta) throughout the entire paper.

Army Project Number
2Q762722A764

Unit Training Standards
and Evaluation

Research Memorandum 78-19

USING STEIN'S ESTIMATOR TO PREDICT UNIVERSE SCORES FROM OBTAINED SCORES

Frederick H. Steinheiser, Jr., and Stephen L. Hirshfeld

Angelo Mirabella, Work Unit Leader

UNIT TRAINING AND EVALUATION SYSTEMS TECHNICAL AREA

August 1978

Submitted as complete and
technically accurate, by
Frank J. Harris
Technical Area Chief

Approved by:

A. H. Birnbaum, Acting Director
Organizations and Systems
Research Laboratory

Joseph Zeidner, Technical Director
(Designate)
U.S. Army Research Institute for
the Behavior and Social Sciences

Research Memorandums are informal reports on technical research problems. Limited distribution is made, primarily to personnel engaged in research for the Army Research Institute.

USING STEIN'S ESTIMATOR TO PREDICT UNIVERSE SCORES FROM OBTAINED SCORES

The purpose of this paper is to introduce and apply a recently developed statistical method for estimating true (population) scores from observed (sample) scores. Provided that at least three scores are available, this method overall will give more accurate true score estimates than the individual maximum likelihood estimates (MLE), regardless of the true abilities of the examinees (Efron & Morris, 1977). The method can be used without knowledge of the (Bayesian) prior distribution, and normality of the true scores being estimated need not be assumed. The theoretical and practical implications of the method extend beyond psychological measurement to the very foundations of statistical inference and have caused some tumult in that discipline during the past decade.

HISTORICAL OVERVIEW

For the Gaussian distribution, the average is the best estimator of the true mean, θ. The average is said to be "unbiased" because no single value of θ is favored over any other value. That is, the expected value of the average, $\bar{x}$, equals the true value of θ, regardless of the value of θ. How many unbiased estimates of θ are there? An infinite number. But, none of them estimates θ perfectly. The expected squared error of estimation for the average is lower than that for any other linear or nonlinear and unbiased function of the data.

A departure from this classical approach assumes that unbiased estimates of θ are not the only methods by which to infer population values. For example, other possible estimates of θ could be the median, $\bar{x}/2$, $2\bar{x}$, the mode, etc. All such estimators can be compared through a risk function, which is the expected value of the squared error for every possible value of θ. Plots of risk functions show that there is no estimator with a risk function that is everywhere lower than the risk function of the average, $\bar{x}$, provided that a single mean is being estimated. But in the more general case, a score is available from each of many examinees who have taken a test, for example, and it is the true score of each examinee that is to be inferred. Thus, the MLE is merely a specific case of the more general situation where the mean scores (θ's) are sought for each examinee.

Theoretical work conducted by Stein (1955) and by James and Stein (1961) concentrated on estimating several unknown means, through methods other than maximum likelihood estimation. The authors assumed that the means are independent of each other and that the goodness of various estimators can be assessed by a risk function: the sum of the expected values of the squared errors of estimation for all of the individual

means. Also, it is not necessary to assume that the means being estimated come from normal distributions. What James and Stein proved is that when three or more means (θ values) are being estimated, it is a less than optimal solution ("inadmissible") to estimate each θ from its own average. That is, estimation rules can be found with smaller total risk regardless of the values of the true means (θ's) for each examinee. As Efron and Morris (1975) express this accomplishment:

> Charles Stein showed that it is possible to make a uniform improvement on the maximum likelihood estimate (MLE) in terms of total squared error risk when estimating several parameters. . . . This achievement leads immediately to a uniform, nontrivial improvement over the least squares (Gauss-Markov) estimators for the parameters in the usual formulation of the linear model (p. 311).

THE STEIN ESTIMATOR

The following discussion serves as an introduction to the Stein estimator. Assume that we have k parameters $\theta_1, \theta_2, \ldots \theta_k$, $k \geq 3$, and that for each θ_i we observe an independent normal variate x_i with mean $E_{\theta_i} x_i \equiv \theta_i$, and variance $Var_{\theta_i}(x_i) = 1$. Note that each x_i might be the mean of n independent observations $Y_{ij} \sim n(\theta_i, \sigma^2)$. Then $x_i \sim n(\theta_i, \sigma^2/n)$, and a change of scale transforms σ^2/n to the more convenient value of 1. Therefore, the above assumptions often occur as a reduction from more complicated situations to this canonical form.

The primary objective for applying the set of estimation rules is to estimate the unknown vector of means $\vec{\theta}$, $\vec{\theta} \equiv (\theta_1, \theta_2, \ldots \theta_k)$. The performance of an estimation rule is assessed by computing the sum of squared component errors that is the squared error loss for that estimation rule. If $\vec{f} = (f_1, f_2, \ldots f_k)$ is an estimation rule, where f_i is the estimate of θ_i, then the squared error loss $\vec{L}(\vec{\theta}, \vec{f})$ is defined as $\vec{L}(\vec{\theta}, \vec{f}) \equiv \sum_{i=1}^{k} (f_i - \theta_i)^2$.

In the case of the maximum likelihood estimator, or the sample mean, denoted by $\vec{f}^{o}(\vec{X})$, $\vec{f}^{o}(\vec{X}) \equiv (f_1^{o}(\vec{X}), f_2^{o}(\vec{X}), \ldots f_k^{o}(\vec{X})) \equiv (x_1, \ldots x_k)$,

there is a constant risk, $\vec{R}$, with $\vec{R}(\vec{\theta}, \vec{\delta}^0(X)) = \vec{R}(\vec{\theta}, \vec{X}) = E_\theta \sum_{i=1}^{k} (x_i - \theta_i)^2 = k$. (Note that E_θ indicates the expectation over the distribution $x_i|\theta_i \overset{ind}{\sim} N(\theta_i, 1)$ introduced above. Observe that $E_\theta(x_i - \theta_i)^2 = 1$ for each i, i = 1,...k.)

The Stein estimator may be used to estimate θ. Define the Stein estimator, $\vec{\delta}^1(\vec{X}) \equiv (\delta_1^1(\vec{X}), \delta_2^1(\vec{X}), \ldots \delta_k^1(\vec{X})$, $k \geq 3$, as follows:

$$\delta_i^1(\vec{X}) \equiv \mu_i + \left(1 - \frac{(k-2)}{S}\right)(x_i - \mu_i),$$

where $\vec{\mu} \equiv (\mu_i, \ldots \mu_k)$ represents an initial guess at the true mean, $\vec{\theta}$, and S is defined by $S \equiv \Sigma(x_j - \mu_j)^2$. This estimator thus has risk

$$\vec{R}(\vec{\theta}, \vec{\delta}^1(\vec{X})) \equiv E_\theta \sum_{i=1}^{k} (\delta_i^1(\vec{X}) - \theta_i)^2 \leq k - \frac{(k-2)^2}{k - 2 + \sum_{i=1}^{k} (\theta_i - \mu_i)^2} < k$$

for all $\vec{\theta}$. If $\theta_i = \mu_i$ for all i, the risk is 2, which compares quite favorably to k obtained for the sample mean. In any event, the risk for the Stein estimator is less than that for the maximum likelihood estimator. A discussion of how the risk for the Stein estimator was obtained is presented in the last section of this paper.

The Stein estimator has a very natural interpretation in an empirical Bayes context. If the θ_i themselves are a sample from a prior distribution, $\theta_i \overset{ind}{\sim} N(\mu_i, \tau^2)$, i = 1...k, then the Bayes estimate of θ_i is the *a posteriori* mean of θ_i given the data, and $\delta_i^*(x_i)$ is defined by

$$\delta_i^*(x_i) = E\theta_i|x_i = \mu_i + \left(1 - \frac{1}{1+\tau^2}\right)(x_i - \mu_i).$$

In the empirical Bayes situation, τ^2 is unknown, but it can be estimated because marginally the x_i are independently normal with means μ_i and $S = \sum_{j=1}^{k} (x_j - \mu_j)^2 \quad (1 + \tau^2)\, \chi_k^2$, where χ_k^2 is a chi-square distribution with k degrees of freedom. Given that $k \geq 3$, the unbiased estimate

$$E\,\frac{(k-2)}{S} = \frac{1}{1+\tau^2} \text{ is available.}$$

Substituting $\frac{k-2}{S}$ for the unknown $\frac{1}{1+\tau^2}$ in the Bayes estimate $\int_i^*$ results in $\mu_i + \left(1 - \frac{k-2}{S}\right)(x_i - \mu_i)$, which is the Stein estimator.

Predicting Batting Averages Using the Stein Estimator

The following example is adapted from Efron and Morris (1975, 1977). Batting averages for major league baseball players, based upon their first 45 times at bat, were obtained. The objective was to predict each player's batting average for the remainder of the season. A cutoff after the first 45 times at bat was chosen because that number was large enough to insure a satisfactory approximation to the binomial distribution by the normal distribution and because the vast majority of "at bats" for the season would be estimated. The model assumes that hits occur according to a binomial distribution with independence between players. (Requiring the same number of trials for all players, n = 45, assures equal variances; however, the Stein estimator can also be used when variances are unequal. See Efron and Morris, 1975.)

Let Y_i be the batting average of player i, i = 1,...k (k = 12) after the first 45 times at bat. Assume that $nY_i \overset{ind}{\sim} Bin(n, p_i)$, i = 1,...12, where p_i is the true season batting average, i.e., $EY_i = p_i$.

Because the variance of Y_i depends upon the mean, the arc-sin transformation for stabilizing the variance of a binomial distribution is applied: $x_i = f_{45}(Y_i)$, where $f_n(y) = n^{\frac{1}{2}}\text{arc-sin}(2y - 1)$. It can be shown that this transformation results in x_i having nearly unit variance independent of p_i.

The mean θ_i of x_i is given by $\theta_i = f_n(p_i)$. Values of Y_i, p_i, x_i, θ_i, $\int_i^1$, and p_i are listed for players 1 through 12 in Table 1. Batting averages for the first 45 times at bat are listed in column 1. Each player received from 270 to 590 additional "trials" during the season. The batting averages for this seasonal trial number are listed in column 2. Recall that the objective here is to predict each player's column 2 ("true," "population") value using the initially obtained column 1 ("sample") value.

Table 1

Example Using Batting Averages From 12 Players

(1)	(2)	(3)	(4)	(5)	(6)
Y_i	P_i	x_i	θ_i	$\int_i^1$	$\hat{P}_i$
.400	.346	-1.35	-2.10	-2.49	.319
.378	.298	-1.66	-2.79	-2.60	.311
.356	.276	-1.97	-3.11	-2.71	.303
.333	.222	-2.28	-3.96	-2.82	.296
.311	.270	-2.60	-3.20	-2.93	.287
.289	.263	-2.92	-3.32	-3.03	.283
.244	.269	-3.60	-3.23	-3.26	.265
.222	.303	-3.95	-2.71	-3.40	.258
.222	.264	-3.95	-3.30	-3.40	.258
.222	.226	-3.95	-3.89	-3.40	.258
.200	.285	-4.32	-2.98	-3.53	.249
.178	.316	-4.70	-2.53	-3.66	.241

Note. Listing of the MLE Scores and Estimated Universe Scores (columns 1 and 2), Score Transformations (columns 3, 4, and 5), and the Estimated Universe Score from using Stein's estimator (column 6).

The x_i values obtained upon application of the arc-sin transformation to the column 1 batting averages (observed scores) are shown in column 3. Similarly, the θ_i values obtained by applying the arc-sin transformation to the column 2 batting averages are shown in column 4. The Stein estimator values that estimate the θ_i are shown in column 5,

and the values obtained upon retransformation via the arc-sin transformation are given in column 6. The following calculations are examples of the type of computations required. Note that the computations are not at all complex.

For i = 2, $f_n(Y_2) = 45\frac{1}{4}\text{arc-sin}(2(.378) - 1) = -1.66$. Therefore, $x_2 = -1.66$, and is entered in column 3. Similarly, $\theta_2 = f_n(p_2) = 45\frac{1}{4}\text{arc-sin}\,(2(.298) - 1)$. This value is given in column 4. Values for $x_1, x_3, \ldots x_{12}$ and $\theta_1, \theta_3, \ldots \theta_{12}$ are obtained through similar substitutions.

The basic equation for the Stein estimator f_i^1, which allows us to estimate the ith component of θ, is slightly different from the expression introduced previously. We estimate the initial guess $\mu = \mu_i/k$ by $\bar{X} = \Sigma x_i/k$, which shrinks all x_i toward $\bar{X}$. The resulting estimate of the ith component θ_i of θ is given by

$f_i^1(\vec{X}) = \bar{X} + \left(1 - \frac{k - 3}{V}\right)(x_i - \bar{X})$, where $V = \Sigma(x_i - \bar{X})^2$, and $k - 3 = (k - 1) - 2$, because one parameter is estimated.

In the empirical Bayes case, the appropriateness of this formulation follows if $\bar{X}$ is used as the unbiased estimate for μ and $\frac{k - 3}{V}$ as the unbiased estimate for $\frac{1}{1 + \tau^2}$. Therefore, in the case of the example data provided in Table 1,

$$\bar{X} = \Sigma x_i/k = \frac{(-1.35) + \ldots + (-4.70)}{12} = -3.10.$$

The value for $\bar{X}$ may in turn be used to compute V:

$$V = \Sigma(x_i - \bar{X})^2 = (-1.35 - (3.10))^2 + \ldots + (-4.70 - (-3.10))^2 = 13.81.$$

The Stein estimates for $\theta_1 \ldots \theta_{12}$ are derived by substituting the obtained values for $\bar{X}$ and V in the computational equation:

$$\tilde{J}_i^1(\vec{X}) = -3.10 + \left(1 - \frac{12 - 3}{13.81}\right)(x_i - (-3.10)) = .350x_i - 2.02.$$

For example, $\tilde{J}_1^1(\vec{X}) = .350(-1.35) - 2.02 = -2.49$. This value and the values for $\tilde{J}_2^1, \ldots \tilde{J}_{12}^1$ are listed in column 5 of Table 1. These values are finally retransformed to obtain the estimates of the "true score" average for each player in column 6.

The total squared prediction error for $\tilde{J}^1(\vec{X})$ is defined as $(\tilde{J}_1^1 - \theta_1)^2 + \ldots + (\tilde{J}_{12}^1 - \theta_{12})^2 = 4.040$. This value is obtained by subtracting the column 4 value from the column 5 value for each of the 12 players, squaring the differences, and summing.

In the case of the sample mean, X, the total squared prediction error is defined as $\Sigma(x_i - \theta_i)^2 = 15.135$. This value is obtained by subtracting the column 4 value from the column 3 value for each of the 12 players, squaring the differences, and summing.

The adequacy of Stein's estimator relative to the sample mean may be determined by computing their relative efficiency. The efficiency of Stein's estimator relative to the sample mean is defined as

$$\frac{\Sigma(x_i - \theta_i)^2}{\Sigma(\tilde{J}_i^1(\vec{X}) - \theta_i)^2}.$$

In this example, the efficiency is 3.746. In other words, Stein's estimator is nearly four times as "efficient" in predicting "universe" or "true" scores from observed (sample) scores as is the MLE.

Limited Translation Estimators

Stein estimators achieve uniformly lower aggregate risk than the MLE (sample mean), as shown above, but may result in increased risk to individual components of θ. In particular, the Stein estimator may do poorly in estimating θ_i with very large or very small values. Therefore, even though $\tilde{f}^1(\vec{X})$ provides better prediction in the aggregate, one may grossly err with individual components. A desirable compromise would be to have both good aggregate and good individual prediction, where improved individual prediction would occur with minimal, if any, loss in aggregate prediction efficiency. This tradeoff may be achieved by using "limited translation estimators" that reduce individual risk for outlying cases and result in minimal loss in aggregate prediction.

Limited translation estimators are introduced to reduce potentially large mean squared prediction errors associated with individual components. Shrinkage of $\tilde{f}_i^1$ values toward x_i values is accomplished through the estimate $\tilde{f}_i^s$, $0 \leq s \leq 1$, of θ_i. (Here, $\tilde{f}_i^0 = x_i$ and $\tilde{f}_i^1 = \tilde{f}_i^1$). $\tilde{f}_i^s$ is defined to be as close to $\tilde{f}_i^1$ as possible, so long as it does not differ by more than $\left[\frac{(k-1)(k-3)}{kV}\right]^{\frac{1}{2}} \left[D_{k-1}(s)\right]$ standard deviations of x_i from x_i. $D_{k-1}(s)$ is a constant, obtained from a table of limited translation estimators (Efron & Morris, 1972).

Data from the baseball example will now be used to illustrate the application of limited translation estimators. Notice in Table 1 that the first player's season average far exceeds the season averages of the remaining players, an example of an outlying case. In the baseball example, $k = 12$, and V was found to be 13.81. Therefore, by obtaining values for $D_{k-1}(.9)$ and $D_{k-1}(.8)$ from the Efron and Morris (1972) table, it is found that $\tilde{f}_i^{.9}(\vec{X})$ may differ by no more than .75 from x_i and $\tilde{f}_i^{.8}(\vec{X})$ may differ by no more .56 from x_i. In other words, by applying $\tilde{f}_i^{.9}$ it means that if $|\tilde{f}_i^1 - x_i| \leq .75$, then f_i^1 is retained; but if $|\tilde{f}_i^1 - x_i| > .75$, $f_i^{.9}$ is set equal to the value differing from x_i by .75.

Table 2 contains values for the 12 players for p_i, Y_i, $\hat{p}_i^1$, $\hat{p}_i^{.9}$, and $\hat{p}_i^{.8}$. Values for $\hat{p}_i^{.9}$ and $\hat{p}_i^{.8}$ are obtained as follows. Consider the first player, $x_i = -1.35$, and $\tilde{f}_i^1 = -2.49$; therefore $|x_i - f_i^1| = 1.14 > .75$, so $\tilde{f}_i^{.9} = -2.10$, and $|x_i - \tilde{f}_i^1| = 1.14 > .56$. Thus, $\tilde{f}_i^{.8} = -1.91$. These values are retranslated to obtain $\hat{p}_i^{.9} = .346$, and $\hat{p}_i^{.8} = .360$. Notice that $p_i = .346$. Therefore, $\hat{p}_i^{.9}$ provides better prediction for this individual than $\hat{p}_i^1$ or $\hat{p}_i^{.8}$. Also note that $\hat{p}_i^{.8}$ is closer to the p_i value than $\hat{p}_i^1$. All three prediction estimates are closer than the MLE value of $Y_i = .400$. In the case of the second player, though, the $\hat{p}_i^s$ value became farther removed from p_i as the value of s decreases from 1 to .9 to .8. Therefore, the translations are increasing the squared prediction error for that player rather than decreasing it. In the case of the fifth individual, $|\tilde{f}_i^1 - x_i| < .75$ and $|\tilde{f}_i^1 - x_i| < .56$, so the estimated value remains the same under translations s = .9 and s = .8. The estimated value will not change until $|\tilde{f}_i^1 - x_i| < .33$. In this particular example, the translation is increasing the error for many individual components by increasing the difference between the estimate and the true score.

Recall that the efficiency of Stein's estimator, $f^1(\vec{X})$, relative to the sample mean was defined to be

$$\frac{\Sigma(x_i - \theta_i)^2}{\Sigma(\tilde{f}_i^1 - \theta_i)^2} = 3.746.$$

The efficiency of the limited translation estimator $f^{.9}(\vec{X})$ relative to the sample mean is defined to be

$$\frac{\Sigma(x_i - \theta_i)^2}{\Sigma(\tilde{f}_i^{.9} - \theta_i)^2},$$

which equals 3.077. Similarly, for $f^{.8}(\vec{X})$ the relative efficiency equals 2.462. Therefore, in this example $f^1(\vec{X})$ has the greatest efficiency of the three estimators, f^1, $f^{.9}$, and $f^{.8}$.

Table 2

Batting Averages and Their Estimates

p_i	$Y_i(x_i)$	$\hat{P}_i^1(\tilde{J}_i)$	$\hat{P}_i^{.9}(\tilde{J}_i^{.9})^a$	$\hat{P}_i^{.8}(\tilde{J}_i^{.8})^b$	$\lvert x_i - J_i^1 \rvert$
.346	.400 (-1.35)	.319 (-2.49)	.346 (-2.10)	.360 (-1.91)	1.14
.298	.378 (-1.66)	.311 (-2.60)	.324 (-2.41)	.338 (-2.22)	.94
.276	.356 (-1.97)	.303 (-2.71)	.303 (-2.71)	.316 (-2.53)	.74
.222	.333 (-2.28)	.296 (-2.82)	.296 (-2.82)	.296 (-2.82)	.54
.270	.311 (-2.60)	.288 (-2.93)	.288 (-2.93)	.288 (-2.93)	.33
.263	.289 (-2.92)	.282 (-3.03)	.282 (-3.03)	.282 (-3.03)	.11
.269	.244 (-3.60)	.265 (-3.28)	.265 (-3.28)	.265 (-3.28)	.32
.303	.222 (-3.95)	.258 (-3.40)	.258 (-3.40)	.258 (-3.40)	.55
.264	.222 (-3.95)	.258 (-3.40)	.258 (-3.40)	.258 (-3.40)	.55
.226	.222 (-3.95)	.258 (-3.40)	.258 (-3.40)	.258 (-3.40)	.55
.285	.200 (-4.32)	.249 (-3.53)	.246 (-3.57)	.234 (-3.76)	.79
.316	.178 (-4.70)	.241 (-3.66)	.222 (-3.95)	.210 (-4.14)	1.04

a $\left[\frac{(k-1)(k-3)}{kV}\right]^{\frac{1}{2}}$ $D_{k-1}(.9) = .75$

b $\left[\frac{(k-1)(k-3)}{kV}\right]^{\frac{1}{2}}$ $D_{k-1}(.8) = .56$

Relationship Between Aggregate and Individual Component Mean Squared Prediction Errors

Prior information about certain examinees can be used to produce modified estimates of their true or universe scores. In this sense, the estimator functions as an empirical Bayesian prediction model. This procedure is most effectively used when the examinee has highly credible information about specific examinees, which is tantamount to having a high prior probability, in the usual Bayesian sense. As a result, for

these particular examinees, the fit of test scores to "true" scores may be improved considerably by use of a limited translation estimator. However, even though the limited translation estimator yields a lower aggregate squared prediction error for the set of examinees as a whole than does the MLE (sample mean), it may reduce the overall efficiency from that of $f^1(\vec{x})$ by increasing the mean squared prediction errors for other examinees in the population. Therefore, overall efficiency, individual squared prediction error, and prior information available on some examinees must all be considered simultaneously to determine what translation, if any, is to be performed.

If there is uniform prior information about all examinees in the score distribution, it may be best to maximize the aggregate efficiency. If no information about true scores is available, it is impossible to assess which individuals have the greatest squared prediction errors associated with them. Therefore, a good strategy would be to achieve maximal aggregate efficiency.

If prior information is concentrated at the extremes of the score distribution, translations may be applied to bring the predicted score more in line with the type of score that might be expected, based upon prior information. In accomplishing this reduction, however, one must evaluate its effect on aggregate efficiency. First, the individual scores can be adjusted until they are in line with prior expectations, and the resulting aggregate efficiency then evaluated. Or, one can focus on attaining maximum aggregate efficiency and then notice how the scores of examinees for whom prior information is available are influenced by minor translations. A major decision is to determine at what point score-fitting for particular examinees becomes counterproductive or inefficient, because minimal additional improvements are achieved at a high cost to the overall aggregate efficiency.

A case in point is when the "true" score does not fall between the MLE and $f^1(\vec{x})$, but when $f^1(\vec{x})$ falls between the true score and the sample mean. Shrinking the difference between the sample mean and $f^1(\vec{x})$ by application of a limited translation estimator, $f^s(\vec{x})$, actually increases the squared prediction error for that examinee. The reasoning is the same when all prior information on an examinee does not fall between the MLE (sample mean) and $f^1(\vec{x})$.

There are also several methodological considerations in relating obtained and "true" score estimates. Initial trials may underestimate a "true" score if the learning curve has not yet reached asymptote in this number of test trials. Likewise, fatigue from the last group of test items could produce an underestimate of the "true" score.

Many factors need to be considered in relating observed scores and true scores, in applying limited translations, in optimizing good individual and good aggregate prediction, and in using prior information on specific examinees productively. The usefulness of Stein's estimator in behavioral and educational research largely depends upon how well these considerations are addressed.

SUMMARY

The scientific implications and practical applications of the Stein estimator approach for estimating true scores from observed scores are of potentially great importance. The conceptual complexity is not much greater than that required for more conventional regression models. The empirical Bayesian aspect allows the examiner to incorporate his/her own degree of prior information about selected examinees. This approach allows for a more accurate estimation of true scores, with the corollary of using fewer test items to achieve those true score estimates. Efron and Morris (1975) make the point that "there is little penalty for using the rules discussed here because they cannot give large total mean squared error than the MLE. . . ." This assurance may be a sufficient reason for more careful examination of the utility of the Stein estimator and its limited translation estimators as they apply to behavioral and social science research.

REFERENCES

Efron, B., and Morris, C. Limiting the Risk of Bayes and Empirical Bayes Estimators--Part II: The Empirical Bayes Case. Journal of the American Statistical Association, 1972, 67, 130-138.

Efron, B., and Morris, C. Stein's Estimation Rule and Its Competitors--An Empirical Bayes Approach. Journal of the American Statistical Association, 1973, 86, 117-130.

Efron, B., and Morris, C. Data Analysis Using Stein's Estimator and Its Generalizations. Journal of the American Statistical Association, 1975, 70, 311-319.

Efron, B., and Morris, C. Stein's Paradox in Statistics. Scientific American, May 1977, 119-127.

James, W., and Stein, C. Estimation with Quadratic Loss, In Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Vol. I. Berkeley: University of California Press, 1961, pp. 361-379.

Stein, C. Inadmissibility of the Usual Estimator for the Mean of a Multivariate Normal Distribution, In Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, Vol. I. Berkeley: University of California Press, 1955, pp. 197-206.

Stein, C. Confidence Sets for the Mean of a Multivariate Normal Distribution. Journal of the Royal Statistical Society, 1962, 24, Ser. B, 265-296.